Behavior Research Methods
2005, 37 (1), 170-181

Nonlinear, near photo-realistic caricatures using a parametric facial appearance model

STUART J. GIBSON, CHRISTOPHER J. SOLOMON, and ALVARO PALLARES-BEJARANO
University of Kent, Canterbury, England

A mathematical model previously developed for use in computer vision applications is presented as an empirical model for face space. The term *appearance space* is used to distinguish this from previous models. Appearance space is a linear vector space that is dimensionally optimal, enables us to model and describe any human facial appearance, and possesses characteristics that are plausible for the representation of psychological face space. Randomly sampling from a multivariate distribution for a location in appearance space produces entirely plausible faces, and manipulation of a small set of defining parameters enables the automatic generation of photo-realistic caricatures. The appearance space model leads us to the new concept of nonlinear caricatures, and we show that the accepted linear method for caricature is only a special case of a more general paradigm. Nonlinear methods are also viable, and we present examples of photographic quality caricatures, using a number of different transformation functions. Results of a simple experiment are presented that suggest that nonlinear transformations can accurately capture key aspects of the caricature effect. Finally, we discuss the relationship between appearance space, caricature, and facial distinctiveness. On the basis of our new theoretical framework, we suggest an experimental approach that can yield new evidence for the plausibility of face space and its ability to explain processes of recognition.

One of the most favored psychological models for facial recognition is that of *face space*. This framework has been used in attempts to provide a unified account of many aspects of face recognition, including the effects of distinctiveness, caricature, inversion, and race (Byatt & Rhodes 1998; Valentine, 1991), and draws on a wide variety of experimental evidence (Valentine, 2000). The basic paradigm here is to consider any face as a point lying within a multidimensional space, the axes of which correspond to given features or, more generally, descriptors of the human face. In general, a face space may be considered to comprise any number of dimensions—the key requirement being that there must be a sufficient number of degrees of freedom to model any face in the population. Thus, one possible approach would be to assign some dimensions to feature-based descriptors (eyes, nose, mouth, etc.) and other dimensions to configurational aspects (e.g., interocular distance). However, a favored approach is that of principal components analysis (PCA). PCA has received a good deal of attention both in the technical literature as a means of achieving automated face recognition (Moghaddam & Pentland, 1997; Sirovich & Kirby,1987; Turk & Pentland, 1991) and as a plausible model for psychological face space (Bruce, Hancock, & Burton, 1998; Burton, Bruce, & Hancock, 1999; O'Toole, Deffenbacher, Valentin, & Abdi, 1994; Scheuchenpflug, 1999). In the PCA approach, a face is expressed as a linear superposition of *global* facial components, the extension along each dimension or axis of the face space expressing the amount of a given global component that is present in the given face. Valentine (1991) originally proposed two possible (closely related) coding mechanisms in face space. In absolute-based coding (ABC), faces are encoded as absolute values on a set of shared dimensions or axes, whereas in norm-based coding (NBC), faces are represented as deviations from a norm or facial prototype within the face space. Valentine's (1991) original work was general in nature and did not explicitly specify the dimensions of the face space.

Irrespective of the precise nature of the axes in face space, one of the major successes of the multidimensional space framework is the plausible explanation that it provides for the phenomena of facial distinctiveness and caricature. Valentine's (1991) multidimensional space face recognition framework supports both an NBC and an (exemplar-only) ABC model. Thus, according to the NBC model, a face is distinctive if it deviates significantly from a norm or mean position within the face space. This would explain why a caricature, which magnifies this deviation, thereby making the face more distinctive, is effective and can even enhance recognition capacity/identity (Lee & Perrett, 2000). In the ABC model, a face is distinctive if it occupies a region in the face space that has a low exemplar density (i.e., has a low probability density of being occupied) and is not related to its distance from any norm.

The authors thank Tim Valentine and Vivien Moore for interesting discussions and encouraging us to write this article. We also thank Peter Hancock and Michael Lewis, whose comments helped us clarify aspects of this study. Correspondence concerning this article should be addressed to C. J. Solomon, School of Physical Sciences, University of Kent, Canterbury CT2 7NR, England (e-mail: c.j.solomon@ukc. ac.uk).

Note—This article was accepted by the previous editor, Jonathan Vaughan.

According to this basic framework, recognition occurs when a face space representation of a stimulus is matched to a stored representation of a previously encountered face (target). Matching occurs provided the encoded stimulus, which will have some random error, is "closer" to the target than to any other, neighboring face. Although the concept of face space seems psychologically plausible and is broadly endorsed by the work of many in the psychological community, its nature has yet to be made sufficiently explicit to address many outstanding questions. It has been suggested (Byatt & Rhodes, 1998) that

> Progress towards a more complete understanding of face recognition requires a more detailed specification of the structure of face-space and the perceptual information in faces. Modeling techniques which derive features from the statistical structure of a set of faces may have considerable potential for quantifying the perceptual dimensions of face-space.

This article has two main aims. The first is precisely to introduce such an explicit model based on statistical training of the shape and textural characteristics of a representative sample of faces. The second aim is to show how this explicit model leads to a new perspective on the related issues of caricature and facial distinctiveness.

In the first section, we will outline the proposed model for face space—*appearance space.* Appearance models were originally intended as a tool for automatically locating objects of interest in digital images (Cootes, Edwards, & Taylor, 1998), and their use in computer vision applications has been well documented. Appearance space is a vector space, based on a decomposition of a training sample of faces into principal components, that subtly differs from the PCA spaces considered so far in the psychological literature (e.g., Burton et al., 1999; Hancock, 2000; O'Toole et al., 1994). *Shape* and *texture*[1] are inherently unified in the appearance model representation, defining a minimal set of dimensions (capable of describing any facial appearance) along which *global shape–texture* characteristics lie. We will present a geometric qualitative account of appearance, highlighting the key features of this model and referring the reader to the Appendix for the mathematical details.

In the second section of this article, we will discuss the current understanding and methods for generating photo-quality caricatures and will show how the appearance space representation naturally suggests a more general paradigm for caricatures and offers a simple means to investigate the relationship between caricature, distinctiveness, and recognition. We will give illustrative examples of new, *nonlinear* caricatures that can easily be generated using this model. We will conclude with a brief summary and discussion in the third section.

Appearance Space

From a technical perspective, we may consider images of the human face to exhibit two key properties: *shape* and *texture*. These two properties are interrelated, since the perception of shape in grayscale and color is derived from transitions in the textural content of the image (predominantly, sharp transitions depicting edges). Although the reduction of the complex series of visual cues that take place in face recognition cannot be explained simply in terms of such general properties, shape and texture nonetheless constitute the base components with which the image analyst can and often must operate. So-called *active appearance* models, which combine the shape and the texture in an image into a single set of optimally compact parameters, have been developed in the computer vision community as a promising approach to automated face recognition (Cootes et al., 1998) and for predicting the effects of aging (Lanitis, Taylor, & Cootes, 2002). The same parametric appearance model has also been successfully employed to produce near-photographic-quality facial composites for use in criminal investigation procedures (Gibson, Pallares, & Solomon, 2003).

From a practical perspective, it is instructive to consider the appearance space model as comprising three necessary elements or parts: (1) the *training* or generation of the appearance model from a population sample of faces, (2) the *decomposition* of a given face in digital form into its appearance model parameters, and (3) the *synthesis* of a face from its appearance model parameters (i.e., the reverse of decomposition). The three elements of training, decomposition, and synthesis, enable us to *model* human facial appearance, *reduce* a digital representation of a human face to its most compact (parametric) form, and conversely, to *reconstruct* the facial appearance from its parametric form, respectively. We refer the reader to Cootes et al. (1998) for full mathematical details of active appearance models. To keep this article

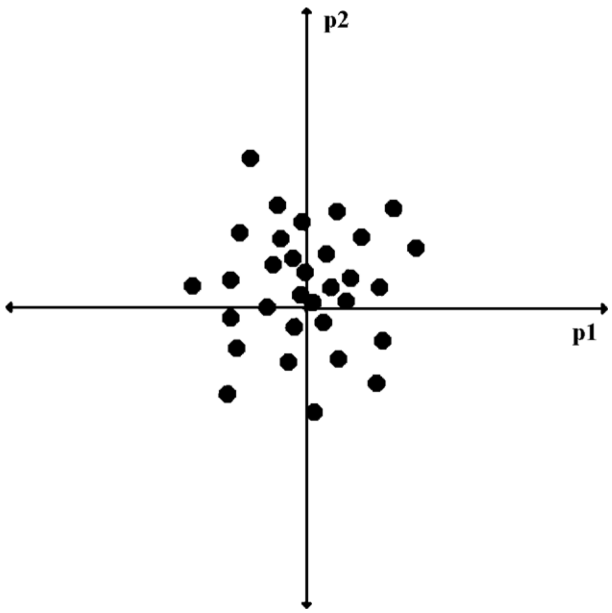

Figure 1. Appearance space comprises a set of orthogonal axes in which faces are represented as points or, equivalently, position vectors within the space. Each axis describes a global shape–texture characteristic. The origin corresponds to the average face.

self-contained but, at the same time, to avoid a lengthy mathematical diversion, we offer a detailed prescription of the computational procedure for the three key steps of training, decomposition, and synthesis in appearance models in the Appendix.

Despite the relative mathematical complexity, we can easily summarize the key properties of appearance space.

1. Any face can be described as a (parametric) vector of coefficients, $\mathbf{c} = [c_1\, c_2 \ldots c_{n-1}\, c_n]^T$, providing the extension of the face along each of the appearance space axes.

2. Appearance space is a multidimensional vector space, the axes of which correspond to specific shape–texture facial characteristics/features. As such, the appearance parameters control the amount of each global shape–texture principal component in the face. All the dimensions in face space represent *commensurate* quantities that describe characteristics of the face as a whole. Such a representation is mathematically and intuitively satisfying.

3. Appearance space is an optimally compact space. The combined linear PCA on shape and texture ensures that the resulting matrix of appearance parameters is optimally compact in the linear least-squares sense. Thus, a representative training sample will enable us to reconstruct both in- and out-of-sample images, to within a given least-squares error, using a minimum number of dimensions in appearance space.

4. The distribution of the appearance model parameters is independent, multivariate normal. The axes of appearance space are labeled according to the amount of facial variation that they explain in order of decreasing variance. For instance, the variance, ${}^2_k$, associated with the kth axis, is greater than the variance associated with the kth+1 axis; hence ${}^2_{k-1} > {}^2_k > {}^2_{k+1}$.

A geometric representation of appearance space is given in Figure 1. Note that although the diagram depicts a two-dimensional vector space, this is a simplification. In reality, there are typically 30–60 useful dimensions in appearance space.

Randomly sampling appearance space for plausible faces. An arbitrary face can be represented as a point in appearance space specified by a position vector from the origin to that point. Since the distribution of each of the components of the vector are *independent* and *normally distributed*, it follows that a "random" face can be generated by randomly sampling a vector of appearance parameters from the normal distribution. The resulting facial image can be reconstructed from the appearance vector according to Equation A1 in the Appendix. In Figure 2, we show some examples of faces generated in this way. It can be seen that although these faces are novel (i.e., do not belong to a real person), they are completely plausible in appearance. In this sense, appearance space satisfies one important criterion for a reasonable heuristic model of psychological face space. It is appropriate at this stage to make two further points.

1. Previous work (Burton, 1994; Burton et al., 1999; Ellis, Burton, Young, & Flude, 1997) has offered evidence to suggest that faces are parameterized in some way. Further support for a PCA-type representation has been of-

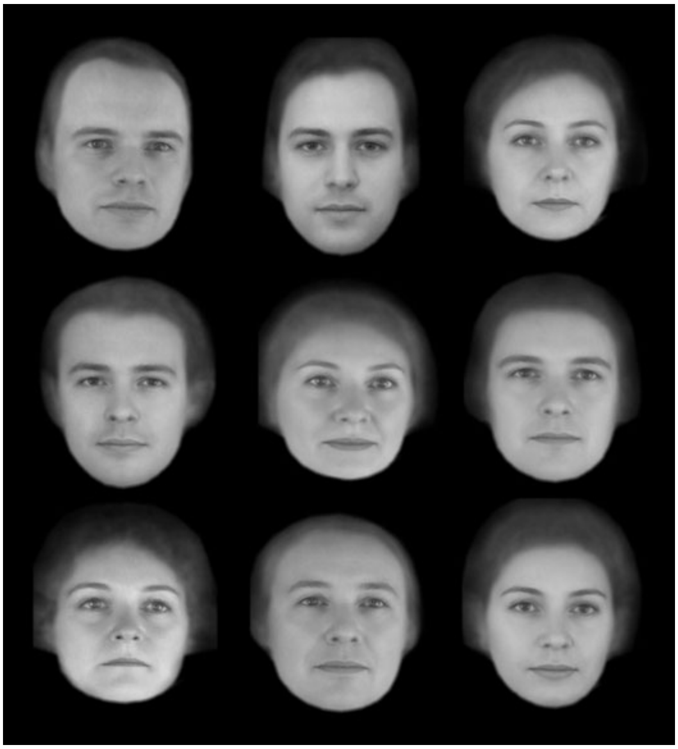

Figure 2. Examples of randomly generated but plausible faces.

fered by Hancock (Hancock, Bruce, & Burton, 1996), who demonstrated that distinctive faces had extreme (i.e., relatively improbable) PCA coefficients, and Burton (Burton et al., 1999) has proposed PCA as the perceptual front-end to the IAC model that has been partly successful in unifying perceptual and cognitive aspects of face recognition. In line with these authors, it is not our intention to suggest that we perform an appearance space decomposition when we recognize a face but only that some of the information delivered by our model shares in common some of the information used in human facial recognition. With the appearance space model, we point to (1) the increased efficiency of the appearance model coding over straight PCA and (2) the identical nature of the axes or dimensions of appearance space as strong indicators of plausibility.

2. The treatment of shape and texture as independent quantities (and hence, the generation of independent models; see, e.g., Hancock, 2000) is, strictly speaking, incorrect. Shape and texture are correlated, and only the appearance space model guarantees a plausible appearance.

In summary, the holistic and efficient encoding of shape and texture is intuitively and mathematically appealing. As such, we may view the appearance model to be a refinement on early work proposing PCA as an heuristic model of face space.

Caricature, Distinctiveness, and Identity

Appearance space is a specific and precise case of the general paradigm proposed by Valentine (1991, 2000) in which each dimension of the face space assigns the amount of a global shape–texture component. It is natural to ask whether the appearance model is consistent with the multidimensional space framework for face recognition and with established recognition phenomena. Accordingly, we will now turn to the question of facial distinctiveness and caricature.

Caricatures exaggerate the distinctive aspects that individuate a particular face, and the study of caricature has important implications for the understanding of facial recognition. The artist creates a caricature by identifying features in the face that deviate from an established norm and exaggerates that difference but, of course, does not attempt to quantify this in precise numeric terms. A substantial body of scientific work (Benson & Perrett, 1991; Rhodes, Brennan, & Carey, 1987) has now shown that the distortions produced by computer-generated caricatures not only do not degrade recognition accuracy but may even enhance it (Lee & Perrett, 2000). Similarly, anticaricatures, in which the deviations from the norm are reduced, are reliably associated with poor recognition performance.

In the following sections, we will begin by outlining the currently accepted method for achieving caricature (Benson & Perrett, 1991, 1993) and then will suggest a more general caricature transformation, showing how the appearance space model offers a new perspective on the related issues of caricature and distinctiveness. Finally, we will suggest how the new forms of caricature can be used experimentally, providing new information on the plausibility of a multidimensional psychological face space and its ability to explain the relation between caricature, facial distinctiveness, and recognition.

Current approach to caricature. Brennan (1985) is generally attributed with the first automatic caricature generator. This worked on line drawings represented as point configurations. The generation of near photo-realistic caricatures was first developed by Benson and Perrett (1991), and subsequent authors have followed their basic approach (Burt & Perrett, 1995; Perrett, May, & Yoshikawa, 1994). Benson and Perrett (1991) considered the shape and texture characteristics of the images separately. In order to understand clearly their method for computer generation of

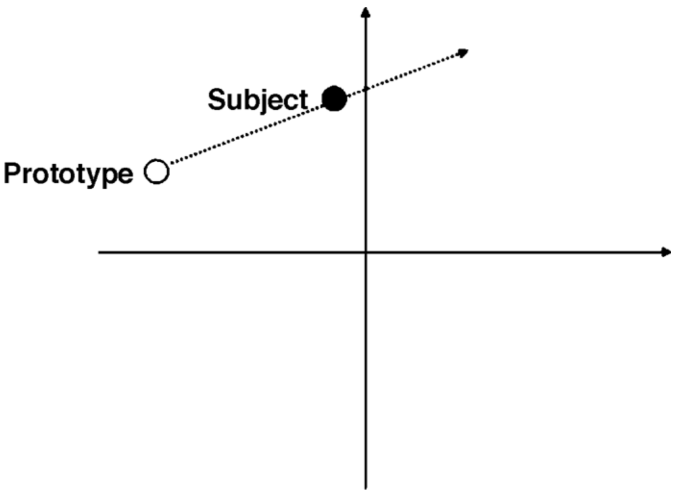

Figure 3. Schematic depicting the approach to uniform linear caricature (as per Benson & Perrett, 1991). The difference vectors between the veridical and the prototype for both shape and texture are simply extended in length.

caricatures and its relation to that proposed herein, we will describe their approach in detail.

1. Facial landmarks corresponding to key points are identified on each image in the sample. The corresponding set of (scaled and aligned) Cartesian coordinates $\{x_i, y_i\}$ for each image constitutes a shape vector for the given face, which we denote by **S**. The mean shape vector over the sample, $\overline{\mathbf{S}}$ is calculated.

2. Each image is warped to the mean shape vector of the entire sample by standard Delaunay triangulation methods. We refer to such warped images as the *shape-normalized texture patterns*, in which the color values are stored as the elements of a vector **X**. The average of the shape-normalized texture patterns is termed the *facial prototype*, $\overline{\mathbf{X}}$.

3. To generate a caricature of any face with texture T and shape X, we (1) calculate the texture difference vector $\mathbf{T} = \mathbf{T} - \overline{\mathbf{T}}$ between the shape-normalized texture pattern and the texture of the facial prototype, (2) calculate the shape difference vector $\mathbf{X} = \mathbf{X} - \overline{\mathbf{X}}$ between the shape vector of the face and the mean shape vector corresponding to the facial prototype, (3) add some linear multiple of the texture and shape difference vectors, $\mathbf{T} \quad \mathbf{T} + a \quad \mathbf{T}$ and $\mathbf{X} \quad \mathbf{X} + b \quad \mathbf{X}$, where a and b are the boost parameters, and (4) finally, warp the texture map $\mathbf{T}$ to the required shape $\mathbf{X}$.

The procedure for shape is represented diagrammatically in Figure 3. A similar diagram may be used to represent the procedure for the texture component. Note that the prototype does not necessarily lie at the origin, although in our own work, we have constructed a vector space in which the mean face prototype does lie at the origin.

A key point to note about this method is its linear (in fact, *uniform*) treatment of all local deviations from the prototype. Thus, in Benson and Perrett's (1991) approach, all differences in local pixel values between the stimulus and the prototype *are* multiplied by an identical factor (a for the texture and b for the shape). This model for caricaturing thus effectively gives all directions in face space equal importance. In anticipation of alternative methods, we term the currently accepted model for caricature the *uniform model*.

Generalized interpretation of caricature. There would appear to be no binding a priori reason to assume that this very simple linear (uniform) relationship correctly models the key underlying principle of the caricature effect. Indeed, a plausible intuitive explanation of our ability to quickly recognize even gross caricatures might be that *recognition is associated primarily with those components or features in a face that are significantly different from the prototype* and that components or features that are close to the norm are not so important. Thus, those features or descriptors that significantly deviate from the norm may be the key to identity (and thus, recognition), and these features/descriptors should be preferentially boosted. Indeed, such an approach seems closer to what the skilled caricature artist does. In the next section, we will present a more general model of caricature (within which the uniform model is simply a special case) and will suggest that other mathematical forms may be appropriate in creating caricatures that maintain and even enhance recognition capacity.

Caricatures in appearance space. In appearance space, all information on facial appearance is expressed in a vector of shape–texture parameters—namely, the appearance vector $\mathbf{c} = [c_1\ c_2\ \ldots\ c_N]^T$. Recall that the appearance parameters are distributed as independent normal variations. It thus follows that the likelihood of a given face's occurring in the population can be measured simply as the scaled distance from the origin. Specifically, the log of the probability density for an appearance vector $\mathbf{c} = [c_1\ c_2\ \ldots\ c_N]^T$ is given by $-\log p = L$, where

$$L = {}_{i=1}^{N} \frac{c_i^2}{{}_i^2}$$

and ${}_i^2$ is the variance associated with the ith axis in the appearance space. Thus, uniform caricaturing moves a face to a region in appearance space where faces are statistically less likely to occur (i.e., of lower exemplar density), but crucially, the shift *is precisely along that original direction* that minimizes the exemplar density. Thus, over a suitable sample of faces, a prototype appearance vector $\bar{\mathbf{c}}$ is easily calculated, and a *uniform* caricature **c** is created by the simple transformation

$$\mathbf{c} = \mathbf{c} + \mathbf{I}(\mathbf{c} - \bar{\mathbf{c}}), \qquad (1)$$

where is a scalar boost parameter and **I** denotes the unit matrix. The reconstruction of the caricatured face from the appearance vector is then governed by applying Equation A1 in the Appendix to **c** , followed by warping to the required shape. If different boosts ${}_S$, ${}_T$ are required for the shape and texture components, this is easily achieved through the decoupled shape and texture parameter vectors $\mathbf{b}_\mathrm{S}$ and $\mathbf{b}_\mathrm{T}$, which are calculable from **c** (see Equation A2).

Uniform caricature transform:

$$\mathbf{b}_\mathrm{S} = \mathbf{b}_\mathrm{S} + {}_\mathrm{S}\mathbf{I}(\mathbf{b}_\mathrm{S} - \overline{\mathbf{b}}_\mathrm{S})$$
$$\mathbf{b}_\mathrm{T} = \mathbf{b}_\mathrm{T} + {}_\mathrm{T}\mathbf{I}(\mathbf{b}_\mathrm{T} - \overline{\mathbf{b}}_\mathrm{T}). \qquad (2)$$

For any given face, it is a simple matter to assess the number of standard deviations by which each parameter deviates from the prototype. It would seem plausible that those global facial components that deviate more drastically from the mean are those largely responsible for encapsulating the distinctive qualities of the individual face. Preferentially boosting these components might be expected to achieve a subtly different kind of caricature that enhances identity-related components (see Figure 4). This suggests that the boost factors ${}_\mathrm{S}$ and ${}_\mathrm{T}$ (simply scalars in uniform caricature) become transforming matrices[2] whose specific value at a point (i.e., for a given appearance parameter) depends on the number of standard

deviations from the prototype at that point. Thus, we propose a more general caricature transform.
General caricature transform:

$$\mathbf{b}_S = \mathbf{b}_S + \boldsymbol{\Gamma}_S \ \ \mathbf{b}_S \quad (\ \mathbf{b}_S = \mathbf{b}_S - \overline{\mathbf{b}}_S)$$
$$\mathbf{b}_T = \mathbf{b}_T + \boldsymbol{\Gamma}_T \ \ \mathbf{b}_T \quad (\ \mathbf{b}_T = \mathbf{b}_T - \overline{\mathbf{b}}_T), \tag{3}$$

where $\boldsymbol{\Gamma}_S$ and $\boldsymbol{\Gamma}_T$ are *diagonal* matrices that weight the individual elements of the difference vector according to their magnitude.

Using the symbol t_k to denote the value of the kth parameter in units of standard deviations from the mean value, we have experimented with the four possible functional mappings defined below and shown in Figure 5.

Transform Matrix:

$$\boldsymbol{\Gamma} = \begin{array}{ccccc} (t_1) & & \cdots & 0 & 0 \\ & (t_2) & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & (t_N) \end{array} \tag{4}$$

Uniform Function:

$$\{t\} = C \qquad \bullet < t < \bullet \tag{5}$$

Step Function:

$$\begin{aligned} \{t\} &= C \qquad |t| \geq t_{\mathrm{MIN}} \\ &= 0 \qquad |t| < t_{\mathrm{MIN}} \\ & \qquad\quad C > 1 \end{aligned} \tag{6}$$

Quadratic Function:

$$\{t\} = at^2 + bt \qquad \bullet < t < \bullet \tag{7}$$

Stretch–Shrink:

$$\begin{aligned} \{t\} &= C \qquad |t| \geq t_{\mathrm{MIN}} \\ &= B \qquad |t| < t_{\mathrm{MIN}} \\ & \qquad\quad C > 1, B < 1. \end{aligned} \tag{8}$$

Clearly, Equations 6–8 give enhanced weighting to appearance components that deviate significantly from the norm and give rise to nonlinear caricature effects. The relative enhancement that is provided can be controlled by the precise choice of constants a, b, B, and C.

Nonlinear caricatures are amenable to a very simple geometric interpretation. Consider that the difference vector between the veridical appearance vector and the norm (which lies at the origin in our model) is $\mathbf{b}$. The caricature is created by applying a transformation matrix, $\boldsymbol{\Gamma}$, to the vector $\mathbf{b}$ and adding this product to the original vector $\mathbf{b}$. The addition of $\boldsymbol{\Gamma} \ \mathbf{b}$ for the generalized nonlinear case defines a *lengthening and rotation* on the veridical vector, whereas the uniform caricature (anticaricature) effects only a lengthening (shortening) of $\mathbf{b}$. In order to make a comparison between the proposed methods for caricature, the vector $\boldsymbol{\Gamma} \ \mathbf{b}$ was scaled to have the same Mahalanobis distance measure in each case (i.e., uniform, step, quadratic, and stretch–shrink methods). The Mahalanobis distance is defined by $[(\boldsymbol{\Gamma} \ \mathbf{b})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Gamma} \ \mathbf{b})]^{1/2}$, where $\boldsymbol{\Lambda}^{-1}$ is the inverse of the covariance matrix constructed from the $\mathbf{b}$ vectors corresponding to the sampled faces. A special property of appearance space is that $\boldsymbol{\Lambda}^{-1}$ is a diagonal matrix, since the distribution of appearance model parameters is independent. The Euclidian distance measure is inadequate for this purpose, because it does not account for the relative importance of each axis (characterized by its associated variance). We are free to make relatively large displacements from the veridical along the most significant axes and retain plausibility. Conversely, only small displacements along the least significant axes are allowed, preventing highly improbable unfacelike caricatures.

Generation of nonlinear caricatures. As a basic test of the methodology and also to visually explore the nature of both conventional (uniform) and nonlinear caricatures, we generated an appearance model according to the procedure outlined in the Appendix on a sample of 71 faces, using 134 landmark points per face (see Figure 6). The sample contained a number of famous faces, 4 of which were caricatured according to the procedure defined in the Generalized Interpretation of Caricature and the Caricatures in Appearance Space sections. Appearance model caricatures using the forms described by Equations 5–8 are shown in Figure 7 for Jackie Chan, Marylin Monroe, George W. Bush, and Mel Gibson, respectively. The nonlinear caricatures have been boosted

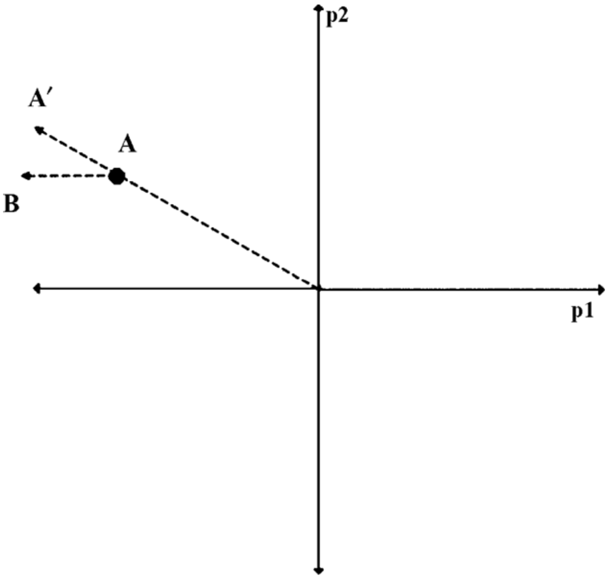

Figure 4. Schematic depicting the idea behind nonlinear caricature. The basic hypothesis is that when the extension along the axes in appearance space is large, these directions should be preferentially weighted. In the three-dimensional representation of face space above, the uniform caricature is defined by *A* . The nonlinear caricature B results because each appearance parameter receives a different weighting as a function of the number of standard deviations from its mean.

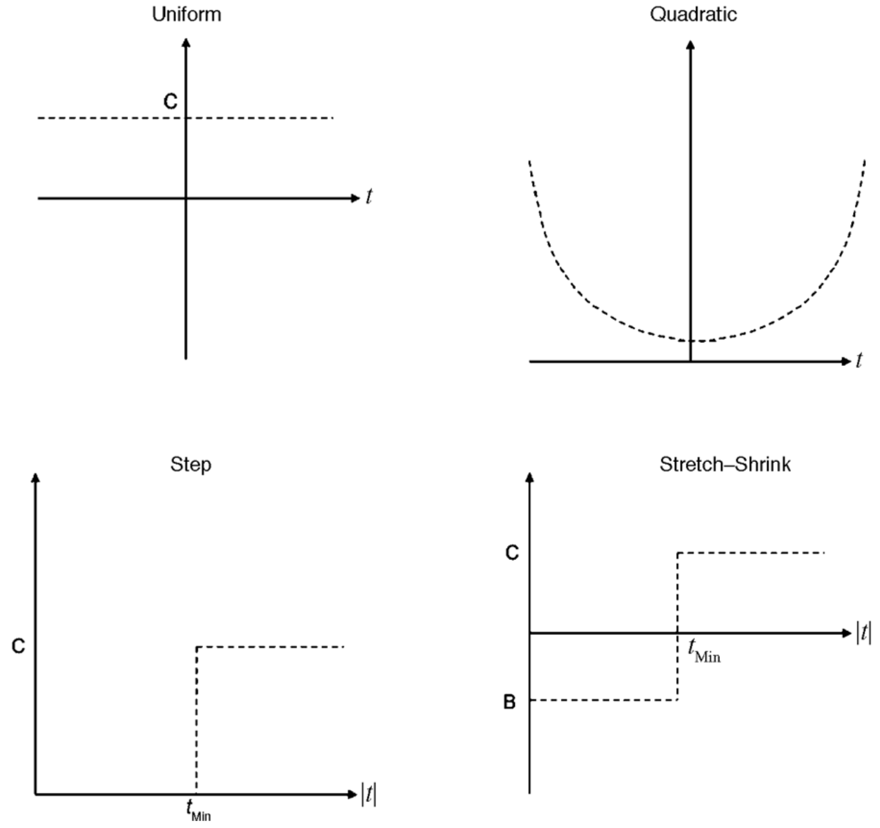

Figure 5. Empirically selected weighting functions for the generation of nonlinear caricatures.

by the same magnitude as the uniform caricatures in accordance with the Mahalanobis distance. Figure 7 is certainly suggestive of the fact that effective caricature can be achieved via nonlinear mappings. The differences in the caricatures produced by the *uniform*, *quadratic*, and *step* functions are subtle but apparent upon careful inspection. Although some images in Figure 7 depict a rather strong degree of caricaturing, the majority of transforms maintain a connection with the basic identity of the subject. The possibility that caricaturing using these methods may enhance identity is suggested; in particular, we note that the step function weighting produces caricatures that are clearly different from the veridical, still appear to maintain basic identity, but do not introduce the rather comic effect characteristic of the uniform transform. Caricatures generated by the *stretch–shrink* function retain some aspects of the identity of the original face but clearly produce significantly different results from the other methods.

Preliminary experiment on nonlinear caricatures. Effective caricatures were achieved by skilled artists long before the mechanism of caricature was subjected to scientific study. Moreover, caricatures of the same subject by different artists often exhibit great variety, suggesting that there is some not inconsiderable flexibility in the caricature method. Given the largely nonlinear behavior of nature, such a nonlinear cognitive model is at least plausible. Whether some nonlinear mapping of the appearance parameter deviations from the prototype better models the cognitive process of recognition and can thereby produce a better-recognized/more-distinctive caricature than the uniform model can must clearly be

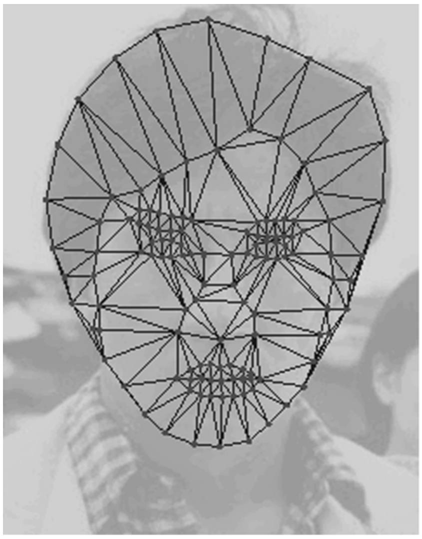

Figure 6. Landmark points used to describe face shape.

Figure 7. Examples of nonlinear and uniform caricatures (Chan, Monroe, Bush, and Gibson). The nonlinear caricatures depicted in this figure have been boosted to the same extent as the uniform caricatures (see the Generation of Nonlinear Caricatures section). For the top row of images, the parameters for each of the four models are as follows: uniform caricature, $\gamma_S = 1$ and $\gamma_T = 0.3$; quadratic caricature, $b = 0$, $\gamma_S = a = 0.3$, and $\gamma_T = 0.3\gamma_S$; step caricature, $t_{MIN} = 1.7sd$, $\gamma_S = 1.5$, and $\gamma_T = 0.3\gamma_S$; stretch–shrink caricature, $t_{MIN} = 2.0sd$, $\gamma_S = 1$, $\gamma_T = 0.3\gamma_S$, $\beta_S = -1$, and $\beta_T = 0.3\beta_S$). The four original images were downloaded from the following web sites: http://www.tbscc.com/users/jherman/President%20Bush.jpg (Bush); http://www.eforu.com/gallery/jackiechan/gallery1.html (Chan); http://www.moviemaze.de/celebs/0014/main.jpg (Gibson); http://www.angelfire.com/ny/marilynmonroegoddess/images/mmhug.jpg (Monroe).

the subject of carefully conducted experiments. Such detailed experiments lie outside the immediate scope of this article. However, a preliminary experiment in which this issue was explored was carried out. Motivated by the results shown in Figure 7, our line of inquiry involved the notion that effective caricatures must generally satisfy two basic criteria—they must maintain identity (i.e., be recognizable as the subjects they are intended to depict), and they must, in the broadest sense, have a comic/humorous appearance—but that each of these aspects may be better achieved by *distinctive* processes.

Thirty randomly selected participants were asked to view four different caricatures (the uniform, step, quadratic, and shrink–stretch mechanisms) of four famous persons (Jackie Chan, Marilyn Monroe, George W. Bush, and Mel Gibson). The images displayed were the four rightmost columns in Figure 7. The true images of the subjects in the leftmost column were *not* displayed to the participants. Prior to showing the caricatures, the participants were asked, Do you know what the subject (e.g., Chan/Monroe/Bush/Gibson) looks like?

For each subject, they were then asked (1) Which of the four caricatures do you find most humorous/comic in appearance? (2) Which of the four corresponds most closely to how you think the subject really looks? In those cases in which the participant did not know what the subject looked like, Question 2 was not asked. The results of this experiment are displayed in Table 1 and summarized graphically in Figure 8.

Inspection of these results suggests that the uniform caricature transform best captures the comic/humorous aspect, whereas the step transform produces a caricature that maintains the closest connection with the real appearance of the subject. The significance of these results was, therefore, assessed using a ^{2} test. Two null hypotheses were examined. (1) *The uniform caricature*

Table 1
Experimental Results for Similarity and Comic Effect Experiments

Subject	Aspect	Uniform	Step	Quadratic	Stretch–Shrink
Chan	Humorous/comic	3	1	9	16
	Similarity to real	2	21	0	1
Monroe	Humorous/comic	19	4	5	2
	Similarity to real	2	11	9	7
Bush	Humorous/comic	21	1	3	4
	Similarity to real	1	22	2	5
Gibson	Humorous/comic	21	4	3	2
	Similarity to real	1	16	10	1
Total	Humorous/comic	64	10	20	24
	Similarity to real	6	70	21	14

mechanism is no more humorous than the others. The calculated ^{2} value was 53.8 with 1 degree of freedom, yielding a probability of much less than 1/1,000 that our experimental data would be obtained if the null hypothesis were true. (2) *The step caricature mechanism is no more effective at capturing realistic appearance than the others*. The calculated ^{2} value was 109.0 with 1 degree of freedom, yielding a probability of much less than 1/1,000 that our experimental data would be obtained if the null hypothesis were true.

In both cases, the results thus provide a highly significant confidence level. Since the distance from the veridical is the same for each type of caricature, it would seem that use of the step transform, in which only the dominant spectral components are boosted, does indeed maintain a better connection with the real appearance of the subject and may be more closely associated with identity.

The subtly different question of which caricature is *best recognized* could be examined more precisely by conducting a careful experiment in which the best recognition may be determined by speed of response to the stimulus (e.g., Lee & Perrett, 1997) or by some other approach. One significance of such an experiment lies in the following argument. If the nonlinear caricatures are best recognized, this will indicate that movement toward a region of minimum exemplar density is *not* the full explanation for why caricatures are effective. This follows from the fact that the distribution of parameters in appearance space is governed by a multivariate normal distribution. Thus, the nonlinear caricature (which adds a vector in a different direction from the direction defined by the veridical image and the prototype) is necessarily in a region of *higher* exemplar density (closer to the origin) than the uniform caricature is. Conversely, if uniform caricatures are best recognized, we obtain an interesting and remarkable result—namely, that the simplest of all models is, in fact, the best one. There would then be even stronger evidence for the multidimensional space hypothesis for a face space in which a caricature is effective because it is distinctive and moves the face to regions in face space where exemplar density is less. Recognition is then easier by virtue of the decrease in the population density of similar faces.

Clearly, the details of such an experiment must be carefully considered. Our intention here is simply to outline the issue in question and the basic approach. Issues such as the nature of the population sample used in the experiment, the specific functional form used to enhance the most deviant components in the nonlinear approach (Equations 6–8 or others), and the matter of how *best recognized* is assessed are all central.

Summary and Discussion

We have introduced and discussed *appearance space* as an heuristic but plausible model of psychological face space. Appearance space is a refinement on separate PCA spaces that unifies the shape and textural aspects of a face, producing a more compact parametric representation of a face. The appearance model representation, in which each component corresponds to a global shape–texture characteristic of the face, has led us to reconsider the mechanism of computer-generated caricatures. A more general cari-

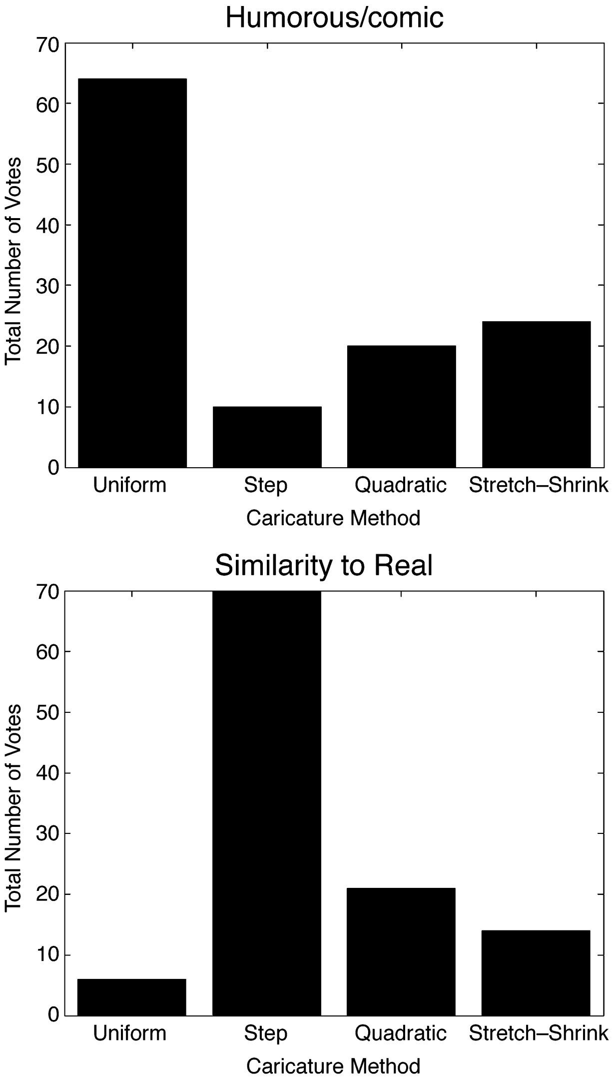

Figure 8. Performance of different methods for caricature summed over the four famous test faces.

cature transformation has been proposed, and specific examples have been presented.

One possible objection to using PCA spaces to represent a face space model is that the exact nature of the appearance space axes depends, to a certain extent, on the sample faces used in the training process. This raises the important question of how to select the training faces in order to construct a face space model. There are two important factors to be considered: the population that is to be modeled and the number of faces to be sampled. In general, the training samples should be drawn from the same population as the faces to be caricatured. (A Caucasian population, 20–60 years of age, was used in the experiment described in this article.) Otherwise, the caricaturing process may emphasize age or ethnicity, rather than aspects of an individual's identity. When the number of training examples is large, the effect of specific sample faces on the model as a whole becomes negligible, and the orientation of the appearance space's axes also becomes fixed for the population of interest. It is also advisable to keep pose and lighting conditions constant where possible. If these precautions are adhered to, the most significant appearance axes will, to a good approximation, be independent of the identity of the training faces. In practice, the least significant appearance axes represent slight, uncontrollable variations in image capture conditions. It is common practice to set a threshold of significance and remove the axes that do not represent aspects of facial identity.

As a consequence of the fixed orientation of the appearance axes, the calculation of a nonlinear transform in an arbitrarily rotated coordinate system (i.e., the axes of the constructed face space are rotated) will produce a different caricature from that obtained in the unrotated system. This raises problems only if one imposes the requirement that the directions of face space must be equivalent. Although this equivalence or isotropic nature is a characteristic of real physical space, it is unnecessarily restrictive to impose this on an abstract vector space. In our case, the calculated axes of appearance space possess unique properties (statistical independence of the parameters); all directions are thus not equivalent, and it is sufficient to define the nonlinear transforms with respect to this specific coordinate system.

The nonlinear approach to caricature transformation that we have proposed is not dependent on having a parametric, appearance space, model of the face. It is certainly possible to define nonlinear mappings on the deviation from the prototype in the separate shape and texture spaces. However, the drawback to working in pixel spaces is the inherent imposition of locality. In other words, we must decide which individual pixels in the texture and which points in the shape model should be preferentially boosted from the prototype values. The *global nature* of each parameter in our appearance space model makes the nonlinear transformation process considerably simpler.

The technique proposed also has potential practical uses. In particular, a major strength of new, emerging methods for the generation of facial composites for use in criminal investigation (Gibson et al., 2003) is that they exploit the powerful human capacity for facial recognition, as distinct from facial recall, which is demonstrably weak in humans. The new caricature method may offer a simple way to further enhance the effectiveness of composites generated by witnesses by enhancing those dominant facial components that are most powerfully associated with identity.

In conclusion, the utility of the appearance space representation and the plausibility of the nonlinear caricature transformation have been demonstrated. Carefully conducted experiments in which such nonlinear caricatures are used have the potential to yield new evidence with regard to the viability of face space and the understanding of facial distinctiveness, and we suggest that these warrant further investigation in the future.

REFERENCES

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Psychology*, **3**, 105-135.

Benson, P. J., & Perrett, D. I. (1993). Extracting prototypical facial images from exemplars. *Perception*, **22**, 257-262.

Brennan, S. E. (1985). The caricature generator. *Leonardo*, **18**, 170-178.

Bruce, V., Hancock, P., & Burton, A. M. (1998). Human face perception and identification. In H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, & T. S. Huang (Eds.), *Face Recognition: From theory to applications* (pp. 51-73). New York: Springer-Verlag.

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London: Series B*, **335**, 137-143.

Burton, A.M. (1994). Learning new faces in an interactive activation and competition model. *Visual Cognition*, **1**, 313-348.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, **23**, 1-31.

Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, **38**, 2455-2468.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998). Active appearance models. In H. Burkhardt & B. Neumann (Eds.), *Proceedings of the European Conference on Computer Vision* (Vol. 2, pp. 484-498). New York: Springer-Verlag.

Ellis, A. W., Burton, A. M., Young, A. M., & Flude, B. M. (1997). Repetition priming between parts and wholes: Tests of a computational model of familiar face recognition. *British Journal of Psychology*, **88**, 579-608.

Gibson, S. J., Pallares, A., & Solomon, C. J. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey & J. A. Banham (Eds.), *Proceedings of the British Machine Vision Conference 2003* (Vol. 1, pp. 221-230). Norwich, U.K.: Page Bros.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B*, **53**, 285-339.

Hancock, P. J. B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments, & Computers*, **32**, 327-333.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition*, **24**, 26-40.

Jolliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Lanitis, A, Taylor, C. J., & Cootes, T. F. (2002). Toward automatic simulation of ageing effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 442-455.

Lee, K. J., & Perrett, D. I. (1997). Presentation time measures of the effects of manipulation in colour space on discrimination of famous faces. *Perception*, **26**, 733-752.

Lee, K. J., & Perrett, D. I. (2000). Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. *Perception*, **29**, 1291-1312.

Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **19**, 696-710.

O'Toole, A. J., Deffenbacher, K. A, Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, **22**, 208-224.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shapes and judgments of female attractiveness. *Nature*, **368**, 239-242.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, **19**, 473-497.

Scheuchenpflug, R. (1999). Predicting face similarity judgements with a computational model of face space. *Acta Psychologica*, **100**, 229-242.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, **4**, 519-524.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience,* **3**, 71-86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, **43A**, 161-204.

Valentine, T. (2000). Face-space models of face recognition. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 83-113). Mahwah, NJ: Erlbaum.

NOTES

1. The term *texture* refers to the pixel values that make up the face image.

2. Technically, they should be considered as matrices, since they effect a rotation of the difference vector between the veridical face and the prototype.

APPENDIX
Facial Appearance Model

Training: The Generation of the Facial Appearance Model

1. The faces in the training set are first hand-marked at a number of control points to form a set of shape model vectors, $\mathbf{S}_i$. The Procrustes-aligned (Goodall, 1991) mean of the shape vectors, $\overline{\mathbf{S}}$, is calculated. We refer to this as the *prototype* shape.

2. A PCA is carried out on the ensemble of aligned shape vectors—that is, we find a linear combination of the shape vectors $\mathbf{P}_\mathrm{S} = (\mathbf{S} - \overline{\mathbf{S}})\mathbf{B}_\mathrm{s}$ that satisfies the required orthogonality relationship $\mathbf{P}_\mathrm{S}^\mathrm{T}\mathbf{P}_\mathrm{S} = {}_\mathrm{S}$ where ${}_\mathrm{S}$ is a diagonal matrix and $\mathbf{P}_\mathrm{S}$ is the matrix containing the principal components. The required diagonalizing matrix $\mathbf{B}_\mathrm{S}$ can be found by standard eigenvector analysis (Jolliffe, 1986).

3. The corresponding texture map vectors $\mathbf{T}_\mathrm{S}$ are warped using standard Delaunay triangulation to the prototype shape. The resulting texture values are referred to as *shape-free* or *shape-normalized* texture maps.

4. A PCA is carried out on the shape-free texture maps. That is to say, we again find a diagonalizing matrix $\mathbf{B}_\mathrm{T}$, such that $\mathbf{P}_\mathrm{T} = (\mathbf{T} - \overline{\mathbf{T}})\mathbf{B}_\mathrm{T}$, with $\mathbf{P}_\mathrm{T}^\mathrm{T}\mathbf{P}_\mathrm{T} = {}_\mathrm{T}$.

5. It is important to recognize that the shape and the texture in a human face are correlated. In the final stage, we combine the separate linear models by decorrelating the shape and the texture. We form a block matrix $\mathbf{B}$,

$$\mathbf{B} = \begin{matrix} \mathbf{W}\mathbf{B}_\mathrm{S} \\ \mathbf{B}_\mathrm{T} \end{matrix},$$

where the upper element of the block contains the eigenvectors that diagonalize the shape covariance and the lower element comprises the eigenvectors that diagonalize the texture (shape-normalized) covariance. The matrix $\mathbf{W}$ is a diagonal matrix of weights that is required to make the shape and the texture parameters, which have different units, commensurate (Cootes et al., 1998). This is achieved by scaling the total variance associated with the texture. In this way, equal weighting is ascribed to the shape and the texture. This process may be described mathematically as

$$\mathbf{W} = r\mathbf{I} = \begin{matrix} r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r \end{matrix}$$

$$r = \frac{{}_{\mathrm{T}i}}{{}_{\mathrm{S}i}},$$

where ${}_{\mathrm{T}i}$ is the variance associated with ith texture principal component and ${}_{\mathrm{S}i}$ is the variance associated with the ith shape principal component.

We apply a further PCA on the columns of $\mathbf{B}$—namely, we seek an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{C} = \mathbf{Q}^\mathrm{T}\mathbf{B},$$

where the columns of $\mathbf{Q}$ are the eigenvectors and $\mathbf{C}$ is the matrix of appearance parameters for the training sample. *The key result here is that each column of* ***C*** *provides a parametric description of the corresponding face in the training sample that is optimally compact in the linear, least-squares sense* (Jolliffe, 1986).

APPENDIX (Continued)

Decomposition of a Face Into Appearance Parameters

Decomposition of a given face into its appearance parameters proceeds by the following stages.

1. The facial landmarks are placed and the Procrustes-aligned shape vector $\mathbf{S}$ of the face is calculated.
2. $\mathbf{S}$ is projected onto the shape principal axes $\mathbf{P}_S$ to yield the *decoupled* shape parameter vector, $\mathbf{b}_S$.
3. The face texture is warped to the prototype or *shape-free* configuration.
4. The shape-free texture map is projected onto the texture principal axes $\mathbf{P}_T$ to yield the decoupled texture appearance parameters.
5. The appearance parameters are calculated using the eigenvector matrix $\mathbf{Q}$:

$$\mathbf{c} = \mathbf{Q}^T\mathbf{b} = \begin{bmatrix}\mathbf{Q}_S^T & \mathbf{Q}_T^T\end{bmatrix} \begin{matrix}\mathbf{W}\mathbf{b}_S \\ \mathbf{b}_T\end{matrix} .$$

Synthesis of Face From Appearance Parameters

The reconstruction of the separate shape and (shape-free) texture vectors of a sample face from its appearance parameters c is calculated through the linearity of the model according to the equations

$$\mathbf{S} = \overline{\mathbf{S}} + \mathbf{P}_S\mathbf{W}_S^{-1}\mathbf{Q}_S\mathbf{c}$$

and

$$\mathbf{T} = \overline{\mathbf{T}} + \mathbf{P}_T\mathbf{Q}_T\mathbf{c}, \tag{A1}$$

where $\overline{\mathbf{S}}$ and $\overline{\mathbf{T}}$ are the mean shape and shape-free textures, $\mathbf{P}_S$ and $\mathbf{P}_T$ are the shape and texture principal components, and $\mathbf{Q}$ is the eigenvector matrix separable into shape and texture block form as

$$\mathbf{Q} = \begin{matrix}\mathbf{Q}_S \\ \mathbf{Q}_T\end{matrix} .$$

The decoupled shape and texture appearance parameters are given by

$$\mathbf{b}_S = \mathbf{W}_S{}^{1}\mathbf{Q}_S\mathbf{c}$$

and

$$\mathbf{b}_T = \mathbf{Q}_T\mathbf{c}. \tag{A2}$$

Warping the shape-free texture to the required shape completes the facial synthesis.

A thorough mathematical description of the appearance model can be found elsewhere (Cootes et al., 1998).

(Manuscript received December 17, 2002;
revision accepted for publication May 13, 2004.)